A NOVEL H.264/AVC BASED MULTI-VIEW VIDEO CODING SCHEME

Akbar Sheikh Akbari, Nishan Canagarajah, David Redmill, Dave Bull and Dimitris Agrafiotis

Department of Electrical and Electronic Engineering, University of Bristol, BS8 1UB, UK

ABSTRACT

This paper investigates extensions of H.264/AVC for compressing multi-view video sequences. The proposed technique re-sorts frames of sequences captured by multiple cameras looking at a person in a scene from different views and generates a single video sequence. The multi-frame referencing property of the H.264/AVC, which enables exploitation of the spatial and temporal redundancy contained in the multi-view sequences, is employed to implement several modes of operation in the proposed coding algorithm. To evaluate the performance of the proposed coding technique at different modes of operations, five multi-view video sequences at different frame rates were coded using the proposed and the simulcast H.264/AVC coding schemes. Experiments show the superior performance of the proposed coding scheme when coding the multi-view sequences at low and up to half of the original frame rates.

Index Terms— Video codecs, multidimensional coding, stereo vision

1. INTRODUCTION

The usage of multi-view image/video will be an important part of highly realistic visual communication environments [1]. However, the price for adding the realism enormously increases the amount of data to be stored or transmitted. A multiview imaging environment consists of an array of cameras which image the world from different positions and viewing angles. This kind of imaging is used in creation of virtual environments, tracking and surveillance applications, medical imaging, entertainment, immersive teleconference, and 3D reconstruction. As the number of camera views increases, the size of the dataset increases linearly. Since all the cameras capture the same world scene, the images/video sequences in the dataset have considerable redundancy. Therefore an efficient compression scheme is vital to exploit the redundancy among the multi-view images/videos to reduce the size of the compressed dataset and realize 3D imaging systems [2], [3].

In recent years, several stereoscopic and multi-view video coding schemes have been proposed [3]-[10]. The main target of these techniques is to efficiently utilize the correlations between the neighboring views along with the spatial and temporal correlation within a single view.

H.264/AVC is the latest development in monoscopic video coding schemes that supports multiple reference frame motion compensated prediction, i.e. more than one prior-coded frame can be used as a reference for motion compensated prediction. This feature of H.264/AVC has made it efficient for coding multi-view video sequences. Several multi-view video codecs have been reported in the literature that use H.264/AVC as a prototype to code multi-view video sequences [4]-[8]. In this case there are some reference frames of neighboring views. Therefore each macroblock prediction could come from either disparity compensated or motion compensated prediction depending on which gives a smaller prediction error.

A H.264/AVC based stereoscopic video codec is reported in [8]. In this coding scheme frames in the left view are just motion compensated while frames in the right views are both motion and disparity compensated. Results indicate significant improvement compared to simulcast coding. In [7], a multi-view H.264/AVC based video codec for coding dynamic light field sequences is proposed. It uses the latest coded frames of all neighboring cameras and the last coded frame of the current camera to predict the current frame of each camera. Authors reported a significant improvement over simulcast coding when dealing with dense light field data-sets although their experiments are based on a part of the views, as original buffering structure of H.264/AVC does not support such a large amount of data.

Bilen et al. [4] proposed another H.264/AVC based multiview video coding scheme. In this codec the buffering structure of H.264/AVC is modified to realize multiple referencing modes that are defined in the codec. In addition, frame numbering and the network abstraction layer unit structure of H.264/AVC are manipulated to provide standard compatibility of the codec in the case of stereoscopic sequences. Experimental results produced using light field sequences were reported in [7]. The experimental results show that for closely located cameras the proposed codec outperforms the simulcast H.264/AVC codec, while for sparsely located cameras their codec improves the coding gain depending on the video characteristics.

In this paper, a new H.264/AVC based multi-view video coding scheme is presented. The proposed coding scheme re-sorts frames of sequences captured by multiple cameras

1-4244-0722-2/07/$20.00 ©2007 IEEE

from a scene and produces a single sequence. Version JM 11.0 of H.264/AVC software is modified in several ways i.e. frame numbering, decoded picture buffer structure and size that enables normal and re-ordered multiple referencing modes, as well as transient operation of those modes prior to the decoded picture buffer being filled. Four modes of operation are implemented within H.264/AVC software to compress the re-sorted sequences. Experimental results were performed on five video conferencing multi-view test sequences at different frame and bit rates. Results indicate that the proposed coding schemes outperform simulcast H.264/AVC at low and up to half of the original frame rates. The rest of the paper is organized as follows: in Section 2 the test sequences are introduced; in Section 3 the proposed coding scheme and its modes of operations are discussed; Section 4 presents experimental results; finally Section 5 concludes the paper.

2. TEST SEQUENCES

Multi-view video conferencing sequences were captured using five cameras in a dedicated studio. Dalsa DS-25-02M30 color cameras that provide 1920×1080 resolution and output up to 30 frames per second at full resolution were used. The cameras were synchronized by an external trigger and the video sequences were stored in a disk. The studio is equipped with a lighting grid to provide controlled lighting conditions and a curtain for background segmentation. All cameras were color calibrated by white-balancing the RGB output with a white reference banner. The cameras were positioned on the circumference of a circle to provide a ring of 5 cameras. The distance between neighboring camera positions was 20 cm. The intrinsic and extrinsic camera parameters were calibrated using the Camera Calibration Toolbox developed in Bristol University.

Five sets of multi-view video conferencing sequences, called Akbar, Damien, Golasa, Lydia and Samira, were captured. Each set contains of five views of 100 frames each at 25 fps. The sequences were first color balanced and gamma corrected. They were then filtered, decimated and clipped to produce CIF and QCIF size sequences. Extrinsic camera parameters were adjusted according to the changes applied to the original sequences to generate CIF and QCIF sequences. To illustrate the nature of the captured datasets, frame 63 of the two side cameras from Golasa sequences are shown in Figure 1.

3. H.264/AVC BASED MULTI-VIEW VIDEO CODING SCHEME

The multi-frame referencing is the key property of the H.264/AVC standard that enables prediction of blocks of a P-frame being coded using a previous coded I-frame or multiple previous coded P-frames. The fact that there are high correlations among different views of a multiview

a) Right camera **b) Left Camera**

Figure 1: Frame 63 of two utmost cameras from Golasa sequences.

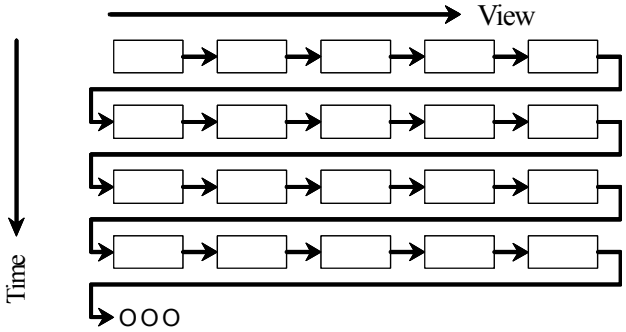


Figure 2: Re-sorted multi-view image sequences.

sequence led to the development of a H.264/AVC based multiview video coding technique with four modes of operation. Hence each macroblock prediction could come from either disparity compensated or motion compensated. Multi-view video sequences are re-sorted to generate a single sequence. Monoscopic H.264/AVC software was modified in several ways to support different frame referencing and reference frame re-ordering. These changes include: Adding camera number of the frame to the DecodedPictureBuffer, adding number of views and the view order to the mbuffer, changing the LevelIDC in the configuration file, which enables the codec supports up to 16 CIF reference frames, filtering the DecodedPictureBuffer in such a way to reference only frames in the reference mode and adding a mechanism to re-order the reference frames. The modified H.264/AVC is then used to code the resulting re-sorted sequence. Figure 2 shows the way that multi-view image sequences are re-sorted to generate a single sequence. In this figure each vertical column represents the image sequence of one view and blocks show captured frame at different time steps. Four different mode of operation of the H.264/AVC based multi-view video coding scheme are explained in the following.

3.1. Mode 1

A block diagram of Mode 1 of operation is shown in Figure 3a, where block number 0 to 15 represent the latest 16 decoded frames in the frame buffer. It can be seen that the

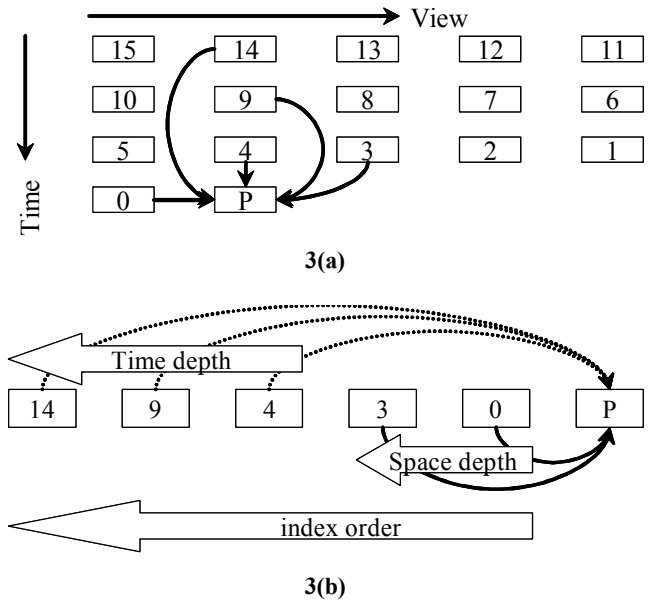


Figure 3: a) Reference frames for Mode 1 and 3, b) Index order for Mode 1.

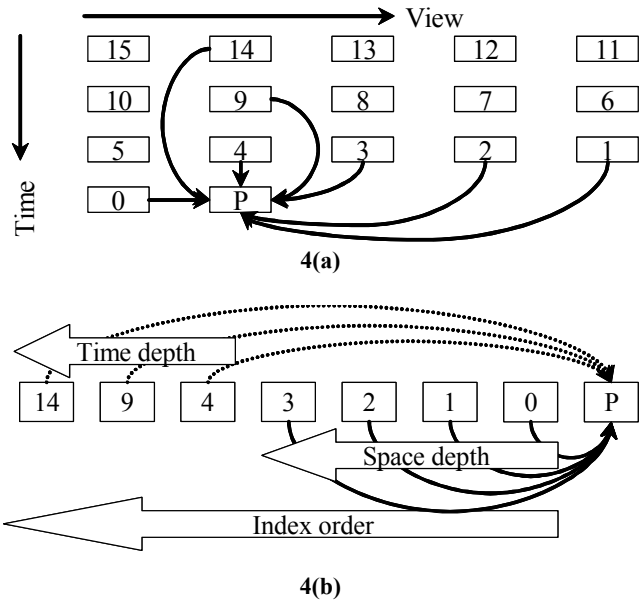


Figure 4: a) Reference frames for Mode 2 and 4, b) Index order for Mode 2.

latest coded frames of two neighboring views and three latest coded frames of the current view are used to predict the frame in the current view. Figure 3b illustrates order of its reference frames in the DecodedPictureBuffer. It shows reference frame that come from neighboring views are assigned lower indices than those are coming from temporal frames. Since H.264/AVC assigns a smaller number of bits to references are made to frames at lower indices, references are made to neighboring views, could be coded with reduced number of bits.

3.2. Mode 2

Figure 4a illustrates the block diagram of Mode 2 of operation, where block number 0 to 15 represent the latest 16 decoded frames in the frame buffer. It employs latest coded frames of all neighboring views and three latest coded frames of the current view to predict the frame in the current view. The order of its reference frames in the DecodedPictureBuffer is illustrated in Figure 4b. It may be seen that reference frame that come from neighboring views are assigned lower indices than those are coming from temporal frames.

3.3. Mode 3 and 4

Mode 3 and 4 of operation are the same as mode 1 and 2 apart from the order of their reference frames in the DecodedPictureBuffer. The order of time depth and space depth reference frames for Mode 3 and 4 are swapped. Hence reference frames that come from temporal frames are assigned lower indices than those are coming from neighboring views. It implies that references are made to temporal frames could be coded with reduced number of bits.

4. EXPERIMENTAL RESULTS

In order to evaluate the performance of the proposed coding schemes, five sets of multiview video conferencing sequences, introduced in Section 2, were used. Each set includes five camera views and every view contains 100 frames in YUV format and CIF size captured at 25 fps. These sequences were temporally down-sampled by frame skipping to generate sequences at 5, 12.5 and 25 fps. Resulting datasets at three different frame rates were then coded using four modes of operation of the proposed coding scheme and simulcast coding technique at different bitrates. The simulcast coding is achieved by coding each view sequence separately using H.264/AVC standard. The quality of the encoded sequences was measured by the average PSNR of their frames. The PSNR of each frame is calculated by $10\log(255\text{^}2/mse)$ where mse is the mean square error of the luminance component of the reconstructed frame. The PSNR measurements for the Y domain of the encoded Akbar sequences using the four modes of operation of the proposed coding scheme and the simulcast coding technique at 5fps and 25 fps and different bitrates are shown in figure 5a and 5b, respectively. Results for coding at 12.5 fps are in between the results presented for coding at 5 and 25fps. From Figure 5a it can be seen that: a) the performance of Modes 2 and 4 of the proposed codec, with respect to the PSNR metric, is significantly higher than the simulcast coding at all bitrates, as each view can be predicted from all other views and three temporal frames. b)

Modes 1 and 3 of the proposed coding scheme give superior results than simulcast coding algorithm at higher bitrates while simulcast coding outperforms them at lower bitrates, since each view may be predicted only from two neighboring views, which do not cover whole object sides, and three temporal views; c) re-ordering the reference frames gives little improvement in the performance of the multi-view video codec, H.264/AVC assigns a smaller number of bits to references made to frames at lower indices but rate distortion optimization significantly reduces its effect. From Figure 5b, which shows the performance of the codecs at original frame rates, it can be seen that: a) the simulcast coding out performs the proposed coding schemes at all bitrates, as temporal correlation within a view is significantly higher than spatial correlation between views; b) modes 2 and 4 of the proposed codec gives higher performance than mode 1 and 3, as they benefit from correlation among all views rather than two neighboring views, where all modes assign more bits to address the reference frames than simulcast. Four other data sets were also coded using the proposed and simulcast coding schemes. Their results confirmed the results presented for the Akbar sequences in this paper.

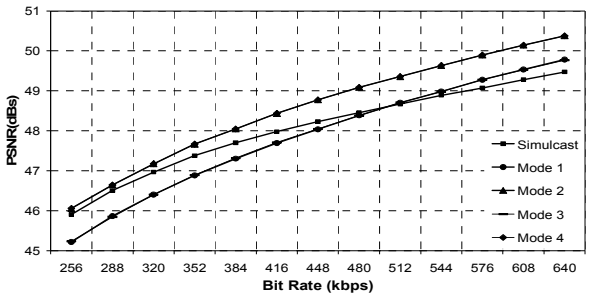
Results indicate the performance of the multi-view video codec depends on the frame rate of the sequences. At low frame rate spatial correlation dominates the temporal correlation, therefore multi-view codec gives higher performance than simulcast codec. Having higher number of reference frames in multi-view codec causes reduction in its performance at high frame rates as they need bits to be addressed. Reference frame re-ordering has little effect on the performance of the multi view codec.

5. CONCLUSION

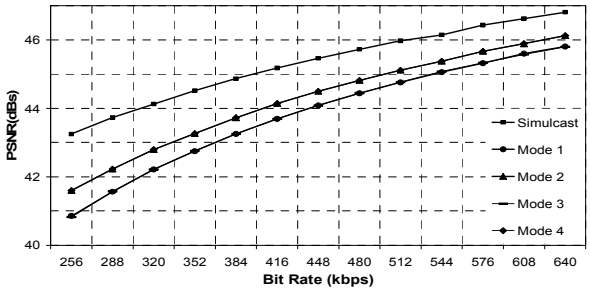
In this paper, a new H.264/AVC based multi-view video coding scheme with four modes of operation was presented. It re-sorts the frames of multi-view sequences and generates a single sequence. H.264/AVC software was developed and four modes of operation were implemented to efficiently extract the spatial and temporal correlations among different views. Five sets of multi-view video conferencing data were generated and used to evaluate the performance of the proposed video codec. Experimental results have shown that the proposed coding scheme outperforms the simulcast H.264/AVC at low and up to half of the original frame rates.

6. AKNOWLEDGEMENT

This work was supported by EPSRC under grant number GR/T11883/01.



5(a) Results for 5fps.



5(b) Results for 25fps.

Figure 5: Average PSNR values for Y domain of Akbar's multi-view sequences at different modes of operation and simulcast coding (Curves for Modes 1 and 3 coincide and also for Modes 2 and 4.).

7. REFERENCES

[1] C. Kamisetty and C. V. Jawahar, "Multi-view Image Compression using Algebraic Constraints", *Proceedings of the IEEE Region 10 Conference on Convergent Technologies (TENCON)*, Bangalore, India, pp. 927—931, October 2003.

[2] N. Anantrasirichai, C. N. Canagarajah and D. R. Bull, "Multi-View Image Coding with Wavelet Lifting and In-Band Disparity Compensation", *ICIP*, Geneva, September 2005.

[3] M. E. Lukacs, "Predictive Coding of Multi-Viewpoint Image Sets", *ICASSP*, pp. 521-524, October 1986.

[4] C. Bilen, A. Aksay, and G. B. Akar, "A Multi-View Video Codec Based on H.264", *ICIP*, Atlanta, October 2006.

[5] G. Li, and Y. He, "A Novel Multi-View Video Coding Scheme Based on H.264", *ICICS-PCM*, Singapore, December 2003.

[6] E. Martinian, H. Sun, A. Vetro, and J. Xin, "Extensions of H.264/AVC for Multi-view Video Compression", *ICIP*, Atlanta, October 2006.

[7] U. Fecker, and A. Kaup, "H.264/AVC-Compatible Coding of Dynamic Light Fields Using Transposed Picture Ordering", *EUSIPCO*, Antalya, Turkey, Sep. 2005.

[8] B. Balasubramaniyam, E. Edirisinghe, and H. Bez, "An Extended H.264 CODEC for Stereoscopic Video Coding", *SPIE 2004.*

[9] R. S. Wang, and Y. Wang, "Multi-view video sequence analysis, compression, and virtual viewpoint synthesis", *IEEE Trans. Sys. Video Technology*, vol. 10, pp. 397-410, April 2000.

[10] J. R. Ohm, "Stereo/Multi-view encoding using the MPEG family of standards", in Proc. SPIE, STEREOSCOPIC displays And Virtual Reality Systems VI, vol. 3639, pp. 242-253, Jan. 1999.